

>LGR4 nucleotide sequence (SEQ ID NO:01)

ATGCCGGGGCCCGCTAGGGCTGCTCTGCTTCCTCGCCCTGGGGCTGCTCGGCTCGGCCGGGGCCAGCGGCGGGCGCCGCT
CTCTGCGGGCGCCCTGCAGCTGCGACGGCGACCGTTCGGGTGGACTGCTCCGAAAGGGGTGACGGCCGTACCGGAGGGT
CTCAGCGCCTTCACCCAAGCACTGGATATCAGTATGAACAATATCACCCAGTTACCAGAAGATGCATTTAAGAGTTTCCCA
TTTCTAGAGGAGCTACAACCTGGCTGGTAACGACCTTTCTCTTATCCATCCAAAAGCCTTGTCTGGGCTGAAAGAAGCTCAA
GTCCTAACACTCCAGAATAATCAGTTGAGAACAGTGCCAGTGAAGCCATTACAGGACTGAGTGCTTTGCAGTCTTTACGC
TTAGATGCCAACCATTATTACCTCAGTCCCGGAGGACAGTTTGAAGGGCTTGTCCAGTTACGCCATCTGTGGCTGGATGAC
AACAGCTTGACGGAAGTGCCCGTGCCTCCCTCAGCAACCTGCCAACCTGCAGGCGCTGACCTTGGCTCTCAACAACATC
TCAAGCATCCCTGACTTCGCTTTCACCAACCTTTCAAGCTTGGTGGTTCCTGCATCTGCATAACAATAAAATTTAAAGCCTC
AGTCAACACTGTTTTGATGGACTAGATAACCTGGAAACCTTGGACTTGAATTACAATTACTTGGATGAGTTTCTCAGGC
ATTAAAGCCCTTCCCAGCCTTAAAGAGCTGGGATTTACAGTAATTCTATTTCTGTTATTCCTGATGGAGCATTTGGTGGT
AATCCACTGCTAAGAAGTATTCAATTTGTATGATAATCCTCTCTTTTGTGGGAAGTCCAGATTTTCAACCTGTCTGAT
CTGCATTGCTTAGTCATTCGTGGTGCAGCCTGGTGACAGTGGTTCCTTTTGTGGGAAGTCCAGTCTGACCGAAGTCCATTTGGAGAGTCTA
ACCTTGACAGGGACAAAAATAAGCAGCATACCTGATGATCTGTGCCAAAACCAAAGATGCTGAGGACTCTGGACTTATCT
TATAACAATATAAGAGACCTTCCAAGTTTTAATGGTTGTCGTGCATTGGAAGAAATTTCAATGCAGCGTAATCAAATCTCC
CTAATAAAGGAAAATACTTTTCAAGGCCTAACATCTCTAAGGATTCTAGATCTGAGTAGAAACCTGATCCGTGAAATTCAC
AGTGGAGCTTTTGCAGAGCTTGGGACAATTACTAACCTGGATGTAAGTTTCAATGAATTAACCTTCAATTTCTACGGAAGGC
CTAAATGGGCTCAATCAACTAAAGCTTGTGGGTAACCTCAAGCTGAAAGACGCTTGGCAGCCAGAGACTTTGCTAATCTC
AGGTCTCTATCAGTACCATATGCTTATCAGTGTGTGCATTTTGGGGGTGTGACTCTTTATGCAATTAACACAGAAGAT
AACAGCCCCCAAGAACACAGTGTGACAAAAGAGAAAGGTGCTACAGATGCAGCAAATGTCACCAGCACTGCTGAGAAGCAA
GAACATAGCCAAATAATTATCCACTGTACACCTTCAACAGGTGCTTTCAAGCCCTGTGAATATTTACTGGGAAGCTGGATG
ATTCGCCTTACAGTGTGGTTCATTTTCTGGTGCCTTGTCTTTTCAACCTGCTTGTCTATTTTAAAGTGTGTGCGTCTTGT
TCATCACTGCCTGCCTCCAAACTCTTCATAGGCTTGATTTCTGTGTCTAAGTACTCATGGGCATCTATACTGGCATCCTT
ACTTTTCTTGATGCTGTGTCTTGGGGCCGATTGGCCGAATTTGGCATTGTTGGTGGGAACTGGCAGCGGCTGCAAGGTAGCC
GGGTCTCTGGCAGTCTTCTCCTCAGAGAGCGCTGTATTCCTATTAACACTGGCAGCTGTGGAAAGAAGCGTATTGTCAAAG
GATTTGATGAAACACGGGAAGAGCAGTCACCTCAGACAGTCCAGGTGGCCGCCCTCTTAGCTTTGTCTGGGTGCCGAGTG
GCAGGCTGCTTCCCCCTTTTCCACGGAGGGCAATATTCTGCATCGCCCTTGTGTTTGGCGTTTCTACAGGAGAAACCCCA
TCGTTAGGATTCACTGTGACCTTAGTGCTATTAACTCACTGGCATTTTTACTAATGGCCATTATCTACACTAACTATAC
TGCAACTTAGAGAAGGAGGACCTGTGCGAAAACCTCCAGTCTAGCGTGATTAAGCACGTTGCCTGGCTCATCTTCACAAAC
TGCATCTTCTTCTGCCCTGTTGCATTTTCTCATTTGCACCATTTGATCACGGCAATCTCCATCAGCCCCGAGATAATGAAG
TCTGTTACACTGATATCTTCCCGTTGCCTGCTTGCCTGAATCCGGTCTGTATGTTTTCTTCAACCCAAAGTTTAAAGAA
GACTGGAAGCTACTGAAGCGCGTGTACCAGGAAACACGGATCTGTTTCAGTTTCCATCAGCAGCCAAGGCGGTTGTGGG
GAACAGGATTTCTACTATGACTGTGGCATGTATCCCACTTGACAGGGTAACCTGACTGTCTGTGACTGCTGTGAGTCATTT
CTTTTGACAAAACAGTATCATGCAAACACTTAATAAAATCGCACAGTTGTCTGTATTGACAGCGGCCTCTTGCCAGAGG
CCAGAGGCTACTGGTCTGATTGTGGTACACAGTCAGCCCATTTGACTATGCAGATGAAGAAGATTCTTTGTCTCAGAC
AGCTCTGACCAGGTGCAGGCTGTGGACGAGCCTGCTTCTACCAGAGTCGTGGATTCCCTCTGGTGCCTATGCTTACAAT
CTACAGAGAGTCAGAGACTGA

>LGR4 amino acid sequence (SEQ ID NO:02)

MPGPLGLLCLFALGLLGSAGPSGAAPPLCAAPCSCDGDRRVDCSGKGLTAVPEGLSAFTQALDISMNNITQLPEDAFKSF
FLEELQLAGNDLSLIHPKALSGLKLKVLTLQNNQLRTPSEAIHGLSALQSLRLDANHITSVPEDSFEGLVQLRHLWLD
NSLTEVPVRPLSNLPTLQALTLALNNISSIPDFAFTNLSSLVVLHLHNNKIKSLSQHCFDGLDNLETLDLNNYLDEF
PQAIKALPSLKLGFHSNSISVIPDGAFGGNPLLRTIHLNDNPLSFVGNSAFHNLSDLHCLVIRGASLVQWFPNLTGTVH
LESLTLTGTKISSIPDDLQCNQKMLRTLDSLNNIRDLPSFNGCRALIEISLQRNQISLIKENTFQGLTSLRILDLSRNL
IREIHSGAFAKLGTITNLDVSFNELTSPFTEGLNGLNQLKLVGNFKLKDALAARDFANLRSLSVPYAYQCCAFWGC
DSLCKLNTEDNSPQEHSVTKEKGATDAANVTSTAENEHSQIIHCTPSTGAFKPCEYLLGSWMIRLTWVFIFLVAL
LFNLLVILTVFASCSSLPASKLFIGLISVSNLLMGIYTGILTFDAVSWGRFAEFGIWWETGSGCKVAGSLAVFSSES
AVFLLTLAAVERSVMFAKDLMKHKGKSSHLRQFQVAALLALLGAAGVAGCFPLFHGGQYSASPLCLFPPTGETPSL
GFTVTLVLLNSLAFLLMAIIYTKLYCNLEKEDLSSENSQSSVIKHVAWLIFTNCIFFCPVAFSFAPLITAISPEIMKSV
TLIFFPLPACLNPNVLYVFFNPKFKEWKLLKRRVTRKHGVSQSVSISSQGGCEQDFYDCGMYSHLQGNLTVCDCESFLL
TKPVSKHLIKSHSCPVLTAAASCQRPEAYWSDCGTQSAHSDYADEEDSFVSDSSDQVQACGRACFYQSRGFPLVRYAYNLQ
RVRD

FIG. 1

>Nucleotide sequence of LGR5 (total 2082 nucleotides) (SEQ ID NO:03)

CTACATCTCCATAACAATAGAATCCACTCCCCTGGGAAAGAAATGCTTTGATGGGCTCCACAGCCTAGAGACTTTAGATTTA
AATTACAATAACCTTGATGAATCCCCACTGCAATTAGGACACTCTCCAACCTAAAGGAACTAGGATTTATAGCAACAAT
ATCAGGTCGATACCTGAGAAAGCATTGTAGGCAACCCCTCTCTTATTACAATACATTTCTATGACAATCCCATCCAATTT
GTTGGGAGATCTGCTTTTCAACATTTACCTGAACTAAGAACACTGACTCTGAATGGTGCCTCACAAATAACTGAATTTCTT
GATTTAACTTGAAGCTGCAAACCTGGAGAGTCTGACTTTAACTGGAGCACAGATCTCATCTCTTCTCAAACCGTCTGCAAT
CAGTTACCTAATCTCCAAGTGCTAGATCTGTCTTACAACCTATTAGAAGATTACCCAGTTTTTCAGTCTGCCAAAAGCTT
CAGAAAATTGACCTAAGACATAATGAAATCTACGAAATTAAAGTTGACACTTTCCAGCAGTTGCTTAGCCTCCGATCGCTG
AATTTGGCTTGGAACAAAATTGCTATTATTCACCCCAATGCATTTTCCACTTTGCCATCCCTAATAAAGCTGGACCTATCG
TCCAACCTCCTGTCGCTCTTTCCCTATAACTGGGTACATGGTTTAACTCACTTAAATTAACAGGAAATCATGCCTTACAG
AGCTGGATATCATCTGAAAACCTTTCCAGAACTCAAGGTATAGAAATGCCTTATGCTTACCAGTGCTGTGCAATTTGGAGTG
TGTGAGAATGCCTATAAGATTTCTAATCAATGGAATAAAGGTGACAACAGCAGTATGGACGACCTTCATAAGAAAGATGCT
GGAATGTTTTAGGCTCAAGATGAACGTGACCTTGAAGATTTCTGCTTGACTTTGAGGAAGACCTGAAAGCCCTTCATTCA
GTGCAAGTGTTCACCTTCCCCAGGCCCCCTTCAAACCCCTGTGAACACCTGCTTGATGGCTGGCTGATCAGAATTGGAGTGTGG
ACCATAGCAGTTCTGGCACTTACTTGTAAATGCTTTGGTGACTTCAACAGTTTTTCAGATCCCCCTCTGTACATTTCCCCCATT
AAACTGTTAATTGGGGTCATCGCAGCAGTGAACATGCTCACGGGAGTCTCCAGTGCCGTGCTGGCTGGTGTGGATGCGTTC
ACTTTTGGCAGCTTTGCACGACATGGTGCCTGGTGGGAGAATGGGGTTGGTTGCCATGTCATTGGTTTTTTGTCCATTTTT
GCTTCAGAATCATCTGTTTTCTGCTTACTCTGGCAGCCCTGGAGCGTGGGTTCTCTGTGAAATATTCTGCAAAATTTGAA
ACGAAAGCTCCATTTTCTAGCCTGAAAGTAATCATTTTGGCTCTGTGCCCTGCTGGCCTTGACCATGGCCGCACTTCCCCCTG
CTGGGTGGCAGCAAGTATGGCGCCTCCCCCTCTCTGCCTGCCTTTGGCTTTTGGGGAGCCCAGCACCATGGGCTACATGGTC
GCTCTCATCTTGCTCAATTCCCTTTGCTTCTCTCATGATGACCATTGCCTACACCAAGCTCTACTGCAATTTGGACAAGGGA
GACCTGGAGAATATTTGGGACTGCTCTATGGTAAAACACATTGCCCTGTTGCTCTTACCACACTGCATCTTAACTGCCCT
GTGGCTTTCTTGCTCTTCTCTCTTTAATAAACCTTACATTTATCAGTCTGAAGTAATTAAGTTTATCCTTCTGGTGGTA
GTCCCACTTCTGTCATGTCTCAATCCCCTTCTCTACATCTTGTTCAATCCTCACTTTAAGGAGGATCTGGTGAGCCTGAGA
AAGCAAACCTACGTCTGGACAAGATCAAAACACCCAAGCTTGATGTCAATTAACCTCTGATGATGTGCAAAAACAGTCTGT
GACTCAACTCAAGCCTTGGTAACCTTTACCAGCTCCAGCATCACTTATGACCTGCCTCCAGTTCCGTGCCATCACCAGCT
TATCCAGTGACTGAGAGCTGCCATCTTCTCTGTGGCATTGTGCCATGTCTCTAA

>amino acid sequence of LGR5 (total 693 amino acids) (SEQ ID NO:04)

LHLHNNRIHSLGKKCFDGLHSLETLDLNNLDEFPTAIRTLSNLKELGFHSNNIRSIPEKAFVGNPSLITIHFYDNPIQF
VGRSAFQHLPELRLTLNGASQITEFPDLTGANLESLLTGAQISSLPQTVCNQLPNLQVLDLSYNLLEDLPFSVCQKL
QKIDLRHNEIYEIKVDTFQQLLSRLSLNLAWNKIAIIHPNAFSTLPSLIKLDLSSNLLSSFPITGLHGLTHLKLGTGNHALQ
SWISSENFPELKVIEMPYAYQCCAFGVCENAYKISNQWNKGDNSSMDDLHKKDAGMFQAQDERDLEDFLDFFEDLKALHS
VQCSPSPGPFKPEHLLDGWLIRIGVWTIAVLALTCNALVTSTVFRSPYIISPIKLLIGVIAAVNMLTGVSSAVLAGVDAF
TFGSFARHGAWWENGVGCHVIGFLSIFASESSVFLLLTAAALERGFVSKYSKFETKAPFSSLVKVIILLCALLALTMAAVPL
LGGSKYGASPLCLPLPFGEPSTMGYMLVALILLNSLCFLMMTIAYTKLYCNLDKGDLENIWDCSMVKHIALLLFTNCILNCP
VAFLSFSSLINLTFISPEVIKIFILLVVVPLPACLNPLLYILFNPHFKEDLVSLRKQTYVWTRSKHPSLMSINSDDVEKQSC
DSTQALVTFTSSSITYDLPPSSVPSPAYPVTESCHLSSVAFVPC

FIG. 2

3/8

>Final LGR7 (LGR7-Long variant) full length sequence (2467 nt) (SEQ ID NO:05).

GAAAGGAGGAAGAAAAAGAGGAATGGAAAGAGACAGAGAAAGGAAATGGGAGTGGGAAGGAGGGAGGACTGCTTT
 GTAAGTCTAAGATTGCAGACAGAAATAGCACACAACCACTGTGAGCTGTATGCGATTACAGAAACCAAGACCAAATT
 TTGCTCACTTTTCAATTAATCAGTTGCTCAGATAGAAGGAAATGACATCTGGTTCTGTCTTCTTCTACATCTTAATTTT
 TGGAAATATTTTTCTCATGGGGGTGGACAGGATGTCAAGTGCTCCCTTGGCTATTTCCCTGTGGGAACATCACAA
 AGTGCTTGCCTCAGCTCCTGCACTGTAACGGTGTGGACGACTGCGGGAATCAGGCCGATGAGGACAACTGTGGAGAC
 AACAATGGATGGTCCATGCAATTTGACAAATATTTTGCCAGTTACTACAAAATGACTTCCCAATATCCTTTTGAGGC
 AGAAACACCTGAATGTTTGGTCGGTTCTGTGCCAGTGCAATGTCTTTGCCAAGGTCTGGAGCTTGACTGTGATGAAA
 CCAATTTACGAGCTGTTCCATCGGTTTCTTCAAATGTGACTGCAATGTCACTTCAGTGGAACTTAATAAGAAAAGCTT
 CCTCTGATTGCTTCAAGAATTATCATGATCTTCAGAAGCTGTACCTGCAAAACAATAAGATTACATCCATCTCCAT
 CTATGCTTTTCAGAGGACTGAATAGCCTTACTAACTGTATCTCAGTCATAACAGAATAACCTTCTCTGAAGCCGGGTG
 TTTTGAAGATCTTCACAGACTAGAATGGCTGATAATTGAAGATAATCACCTCAGTCGAATTTCCCCACCAACATTT
 TATGGACTAAATTTCTTATTCTCTTAGTCCTGATGAATAACGTCCTCACCCGTTTACCTGATAAACCTCTCTGTCA
 ACACATGCCAAGACTACATTGGCTGGACCTTGAAGGCAACCATATCCATAATTTAAGAAATTTGACTTTTATTTCCCT
 GCAGTAATTTAACTGTTTTAGTGATGAGGAAAAACAAATTAATCACTTAAATGAAAATACTTTTGACCTCTCCAG
 AAACCTGGATGAATTGGATTTAGGAAGTAATAAGATTGAAAATCTTCCACCGCTTATATTCAAGGACCTGAAGGAGCT
 GTCACAATTGAATCTTTCTATAATCCAATCCAGAAAATTCAAGCAAACCAATTTGATTATCTTGTCAAACCTCAAGT
 CTCTCAGCCTAGAAGGGATTGAAATTTCAAATATCCAACAAAGGATGTTTAGACCTCTTATGAATCTCTCTCACATA
 TATTTTAAGAAATTCAGTACTGTGGGTATGCACCACATGTTTCGCAGCTGTAAACCAAACACTGATGGAATTTTCATC
 TCTAGAGAATCTCTTGGCAAGCATTATTAGAGAGTATTTGTCTGGGTTGTATCTGCAGTTACCTGCTTTTGAAACA
 TTTTGTGCTTTGCATGCGACCTTATATCAGGTCTGAGAACAGCTGTATGCCATGTCAATCATTTCTCTCTGCTGT
 GCCGACTGCTTAATGGGAATATATTTATTCGTGATCGGAGGCTTTGACCTAAAGTTTCGTGGAGAATACAATAAGCA
 TGCGCAGCTGTGGATGGAGAGTACTCATTGTGAGCTTGTAGGATCTTTGGCCATTCTGTCCACAGAAGTATCAGTTT
 TACTGTTAACATTTCTGACATTGGAAAAATACATCTGCATTGTCTATCCTTTTAGATGTGTGAGACCTGGAAAAATGC
 AGAACAATTACAGTTCTGATTCTCATTGGATTACTGGTTTTATAGTGGCTTTTATTCCATTGAGCAATAAGGAATT
 TTTCAAAACTACTATGGCACCAATGGAGTATGCTTCCCTCTTCATTGAGAAGATACAGAAAGTATTGGAGCCCAGA
 TTTATTTCAGTGGCAATTTTTCTTGGTATTAATTTGGCCGATTATCATCATAGTTTTTTCTATGGAAGCATGTTT
 TATAGTGTTCATCAAAGTGCCATAACAGCAACTGAAATACGGAATCAAGTTAAAAAGAGATGATCCTTGCCAAACG
 TTTTTCTTTATAGTATTTACTGATGCATTATGCTGGATACCCATTTTGTAGTGAAATTTCTTTCACTGCTTCAGG
 TAGAATACACAGTACCATAACCTCTTGGGTAGTGATTTTATTCTGCCATTAAACAGTGCTTTGAACCCAATTCTC
 TATACTCTGACCACAAGACCATTGAAAAATGATTGATTCATCGGTTTATGGTATAACTACAGACAAAGAAATCTATGGA
 CAGCAAAGGTCAGAAAACATATGCTCCATCATTCTGCGGTGGAAATGTGGCCACTGCAGGAGATGCCACCTGAGT
 TAATGAAGCCGACCTTTTACATACCCCTGTGAAATGTCACTGATTTCTCAATCAACGAGACTCAATTCCTATTCA
TGA

>Final LGR7 (LGR7-long variant, total 757 amino acids)(SEQ ID NO:06)

MTSGSVFFYILIFGKYFSHGGQDVKCSLGYFPCGNITKCLPQLLHCNGVDDCGNQADEDCGDNNGWSMQFDKYFA
 SYKMTSQYPFEAETPECLVGSVPVQCLCQGLELDCDETNLRAVPSVSSNVTAMSLQWNLIRKLPPDCFKNYHDLQK
 LYLQNNKITSISIIYAFRGLNSLTCLYLSHNRITFLKPGVFEDLHRLEWLIIEDNHLRSRISPTFYGLNSLILLVLMN
 NVLTRLPLDKPLCQHMPRLHWDLEGNHIIHNLRLNLTIFISCSNLTVLVMRKNKINHLNENTFAPLQKLDELGLGSKNIE
 NLPPLIFKDLKELSQLNLSYNPIQKIQANQFDYLVKLKSLSLLEGIEISNIQQRMFRLMNLSHIYFKKFQYCGYAPH
 VRSCKPNTDGISSLENLLASIIQRFVWVVSAVTCFGNIFVICMRPYIRSENKLYAMSIISLCCADCLMGIYLFVIG
 GFDLKFGEYNKHAQLWMESTHCQLVGLAILSTEVSLLLLFTLTLEKYICIVYFRCVRPGKCRITITVLILIIWITG
 FIVAFIPLSNKEFFKNYYGTNGVCFPLHSEDTESIGAQIYSVAIFLGINLAAPFIIIVFSYSGMFYSVHQSAITATEI
 RNQVKKEMILAKRFFFIVFTDALCWIPFVVKFLSLLQVEIPGTITSWVIFILPINSALNPILYTLTTRPFKEMIH
 RFWYNYRQRKSMDSKGQKTYAPSFIIWEMWPLQEMPPELMKPDLFTYPCEMSLISQSTRNLNSYS*

FIG. 3

4/8

>Final LGR7 (LGR7-Short variant) full length sequence (3584 nt)(SEQ ID NO:07)

CTGCTTTGTAAGTCTAAGATTGCAGACAGAAATAGCACACAACCACTGTGAGCTGTATGCGATTTCAGAAACCAAGA
 CCAAATTTTGGCTCACTTTTCAATTAATCAGTTGCTCAGATAGAAGGAAATGACATCTGGTTCTGTCTTCTTCTACATCT
 TAATTTTGGAAAATATTTTCTCATGGGGGTGGACAGGATGTCAAGTGCTCCCTTGGCTATTTCCCTGTGGGAAC
 ATCACAAGTGCTTGCCTCAGCTCCTGCACTGTAACGGTGTGGACGACTGCGGGAATCAGGCCGATGAGGACAACCTG
 TGTGGTGGTTTTGTGCCAGTGCATGTCTTTGCCAGGTCTGGAGCTTGACTGGATGAAACCATTACAGAGTGTTCAT
 CGGTTTCTTCAAATGTGACTGCAATGTCACTTCAGTGGAACTTAATAAGAAAGCTTCTCTCTGATTGCTTCAAGAAT
 TATCATGATCTTCAGAAGCTGGACCTGCAAAACAATAAGATTACATCCATCTCCATCTATGCTTTTCAGAGGACTGAA
 TAGCCTTACTAACTGTATCTCAGTCATAACAGAATAACCTTCTGAAGCCGGGTGTTTTTGAAGATCTTCACAGAC
 TAGAATGGCTGATAATTGAAGATAATCACCTCAGTCGAATTTCCCCACCAACATTTTATGGACTAAATCTCTTATT
 CTCTTAGTCTGTATGAATAACGTCTCACCCGTTTACCTGATAAACCTCTCTGTCAACACATGCCAAGACTACATTG
 GCTGGACCTTGAAGGCAACCATATCCATAATTTAAGAAATTTGACTTTTATTTCTGCAGTAATTTAACTGTTTTAG
 TGATGAGGAAAAACAAAATTAATCACTTAAATGAAAATACTTTTGACCTCTCCAGAACTGGATGAATTGGATTTA
 GGAAGTAATAAGATTGAAAATCTTCCACCGCTTATATTCAAGGACCTGAAGGAGCTGTCACAATTGAATCTTCTCTA
 TAATCCAATCCAGAAAATCAAGCAAACCAATTTGATTATCTTGTCAAATCAAGTCTCTCAGCCTAGAAGGGATTG
 AAATTTCAAATATCCAACAAGGATGTTTAGACCTCTTATGAATCTCTCTCACATATATTTAAGAAATTTCCAGTAC
 TGTGGGTATGCACCACATGTTTCGAGCTGTAACCAAACACTGATGGAATTTTATCTCTAGAGAATCTCTTGGCAAG
 CATTATTTCAGAGAGTATTTGTCTGGGTGTATCTGCAGTTACCTGCTTTGGAACATTTTGTCTATTGTCATGCGAC
 CTTATATCAGGTCTGAGAACAAGCTGTATGCCATGTCAATCATTTCTCTCTGCTGTGCCGACTGCTTAATGGGAATA
 TATTTATTCGTGATCGGAGGCTTTGACCTAAAGTTTCGTGGAGAATACAATAAGCATGCGCAGCTGTGGATGGAGAG
 TACTCATTGTGAGCTTGTAGGATCTTTGGCCATTCTGTCCACAGAAGTATCAGTTTTACTGTTAAACATTTCTGACAT
 TGGAAAATACATCTGCATTGTCTATCCTTTTAGATGTGTGAGACCTGGAAAATGCAGAACAAATTACAGTTCTGATT
 CTCATTGGATTACTGGTTTTATAGTGGCTTTTATTCCATTGAGCAATAAGGAATTTTCAAAAATCTACTATGGCAC
 CAATGGAGTATGCTTCCCTCTTCATTGAGAAGATACAGAAAGTATTGGAGCCAGATTTATTTCAGTGGCAATTTTTC
 TTGGTATTAATTTGGCCGCATTTATCATCATAGTTTTTCTATGGAAGCATGTTTTATAGTGTTCATCAAAGTGCC
 ATAACAGCAACTGAAATACGGAATCAAGTTAAAAAAGAGATGATCCTTGCCAAACGTTTTTCTTTATAGTATTTAC
 TGATGCATTATGCTGGATACCCATTTTTGTAGTGAAATTTCTTTCACTGCTTCAGGTAGAAATACCAGGTACCATAA
 CCTCTTGGGTAGTGATTTTTATTCTGCCATTAACAGTGCTTTGAACCAATTCTCTATACTCTGACCACAAGACCA
 TTTAAAGAAATGATTCATCGGTTTTGGTATAACTACAGACAAAGAAATCTATGGACAGCAAAGGTGAGAAAACATA
 TGCTCCATCATTCTGCTGGGTGGAATGTGGCCACTGCAGGAGATGCCACCTGAGTTAATGAAGCCGGACCTTTTCA
 CATACCCCTGTGAAATGTCACTGATTTCTCAATCAACAGAGACTCAATTCCTATTTCATGACTGACTCTGAAATTCATT
 TCTTCGCAGAGAATCTGTGGGGGTGCTTCATGAGGGATTTACTGGTATGAAATGAATACCACAAAATTAATTTAT
 AATAATAGCTAAGATAAATATTTTACAAGGACATGAGGAAAAATAAAATGACTAATGCTCTTACAAAGGGAAGTAA
 TTATATCAATAATGTATATATATTAGTAGACATTTTGCATAAGAAATTAAGAGAAATCTACTTCAGTAACATTCATT
 CATTTTTCTAACATGCATTTATTGAGTACCCACTACTATGTGCATAGCATTGCAATATAGTCTGGAAGTAGACAGT
 GCAGAACCTTTCAATCTGTAGATAGTGTAAATGACAAAAGACTATACAAAGTCCATCTGCAGTTCTTAGTTTTAAAG
 TAGAGCTTTACCTGTCTATGTGCATCAGCAAGAATCATAGGCACTTTTAAATAAAGGTTTTAAAGTTTTGGAATACTCA
 GTGTATTGTCATCATGAAAATGTCTGACTGTTTCAAAAATAATATTCTGTTTTAAGAATCCATCTTACCTCTCTTT
 AAGTTTCCATACACTTGAGAGCCAAACACAACATATTGTTTACTATAAAAAGATGCTTTGCTAGAACTCAAAAACAGCA
 CTTCTTTTGGCACTTCTCTGCCAGTTTTCTCTTTGCTTTAAATGAACATCATCATATGGAATTGGAATAGGAGAGTA
 TGAGTACGGCAGAGAAGTGGATCAGAAAAACTAGAATGAGGATAAACATTTACATTAGTGGAACTCCTGAAATAAA
 TCCTTGTATTGTGAGTTAACTGATTTTCAACAAGGATGCCAAGACAAAAGGCTTTTCAACAACCGTGCTGTTTTA
 AGAACAGACCTAAGTGGTTTTAATTCACCCACTTTAGATGGGTGAATGTTATGGTGTGTGAAATATCTCAGTAAAGCA
 GTTAAAGGAAAAAGAGCTGGAATGCACTGATTTCAGGAACCTTAATTTTCAGGAAGGAAAGGTCTGTATGTACACATTT
 CACTTTAAGCAGAAAATCTTTCTTCAAGAAATGACTTTACTTTCTCTTTCAGTGCAGCACGTGAGATACTAATCTT
 TTTAACTAGTTGTTCTTCTCTAGTCTCTACGTTATTAGNATTTTTGCTTTTATAATGTGAAACCTTTAAGCAGGAG
 AAGAAAATGTTTTAGATAGTTTTCAAATACNCCAAAATGTTTGCAACACAAAATACTGGAATCNAACCATAATGC
 CCTTATTGAATATATAGTTGTATAGNTTTGTTCTGAAAACCC

>Final LGR7-S ORF (722 amino acids) (SEQ ID NO:08)

MTSGSVFFYILIFGKYFSHGGQDVKCSLGYFPCGNITKCLPQLLHCNGVDDCGNQADEDNVCVVLCQCMSLPGLLEL
 DWMKPFTSVPSVSSNVTAMSLQWNLIRKLPDPDFKNYHDLQKLDLQNNKITSISIIYAFRGLNSLTKLYLSHNRITFL
 KPGVFEDLHRLEWLIIEDNHLRSISPPTFYGLNSLILLVLMNNVLTPLDKPLCQHMPRLHWDLEGNHIIHNLRLNT
 FISCSNLTVLVMRKNKINHNLNENTFAPLQKLDLQNNKIEHNPPLIFKDLKELSQLNSYNPIQKIQAQNFQDYLV
 KLKSLLEGIEISNIQORMFRPLMNLSHIYFKKFQYCGYAPHRVSRCKPNTDGISSLENLLASIIQRFVFWVVSATVC
 FGNIFVICMRPYIRSENKLYAMSIISLCCADCLMGIYLFVIGGFDLKFGEYNKHAQLWMESTHCQLVGLSLAILSTE
 VSVLLLTFLTLLEKYICIVYPPFRVCPGKCRITITVLILIWITGFIIVAFIPLSNKEFFKNYYGTNGVCFPLHSEDTEI
 GAQIYSVAIFLGINLAAFIIVFSYGSFMFYSVHQSAITATEIRNQVKEMILAKRFFFIVFTDALCWIPFVVKFLS
 LLQVEIPGTITSWVIFILPINSALNPILYTLTTRPFKEMIHRFWYNRQRKSMDSKGQKTYAPSFIVWEMWPLQEM
 PPELMKPDFTYPCEMSLISQSTRLNSYS*

FIG. 4

5/8

>Alignment of LGR7-L with LGR7-S

Query=LGR7-L

Sbjct=LGR7-S

```

Query: 1  MTSGSVFFYILIFGKYFSHGGGQDVKCSLGYFPCGNITKCLPQLLHCNGVDDCGNQADED 60
          MTSGSVFFYILIFGKYFSHGGGQDVKCSLGYFPCGNITKCLPQLLHCNGVDDCGNQADED
Sbjct: 1  MTSGSVFFYILIFGKYFSHGGGQDVKCSLGYFPCGNITKCLPQLLHCNGVDDCGNQADED 60

Query: 61  NCGDNNGWSMQFDKYFASYKMTSQYPFEAETPECLVGSVPVQCLCQ---GLELDCDETN 117
          NC                      V V C C      GLELD  +
Sbjct: 61  NC-----VVVLCQCMSLPGLELDWMKP- 82

Query: 118  LRAVPSVSSNVTAMSLQWNLIRKLPPDCFKNYHDLQKLYLQNNKITSISIIYAFRGLNSLT 177
          +VPSVSSNVTAMSLQWNLIRKLPPDCFKNYHDLQKL  LQNNKITSISIIYAFRGLNSLT
Sbjct: 83  FTSVPSVSSNVTAMSLQWNLIRKLPPDCFKNYHDLQKLDLQNNKITSISIIYAFRGLNSLT 142

Query: 178  KLYLSHNRITFLKPGVFEDLHRLEWLIIEDNHLSRISPPTFYGLNSLILLVLMNNVLTREL 237
          KLYLSHNRITFLKPGVFEDLHRLEWLIIEDNHLSRISPPTFYGLNSLILLVLMNNVLTREL
Sbjct: 143  KLYLSHNRITFLKPGVFEDLHRLEWLIIEDNHLSRISPPTFYGLNSLILLVLMNNVLTREL 202

Query: 238  PDKPLCQHMPRLHWLDLEGNIHNLRLNLTIFISCSNLTVLVMRKNKINHLNENTFAPLQKL 297
          PDKPLCQHMPRLHWLDLEGNIHNLRLNLTIFISCSNLTVLVMRKNKINHLNENTFAPLQKL
Sbjct: 203  PDKPLCQHMPRLHWLDLEGNIHNLRLNLTIFISCSNLTVLVMRKNKINHLNENTFAPLQKL 262

Query: 298  DELDLGSNKIENLPPLIFKDLKELSQLNLSYNPIQKIQANQFDYLVKLKSLSLLEGIEISN 357
          DELDLGSNKIENLPPLIFKDLKELSQLNLSYNPIQKIQANQFDYLVKLKSLSLLEGIEISN
Sbjct: 263  DELDLGSNKIENLPPLIFKDLKELSQLNLSYNPIQKIQANQFDYLVKLKSLSLLEGIEISN 322

Query: 358  IQQRMFRPLMNLSHIYFKKFQYCGYAPHVRSCKPNTDGISSLENLLASIIQRFVFWVWSA 417
          IQQRMFRPLMNLSHIYFKKFQYCGYAPHVRSCKPNTDGISSLENLLASIIQRFVFWVWSA
Sbjct: 323  IQQRMFRPLMNLSHIYFKKFQYCGYAPHVRSCKPNTDGISSLENLLASIIQRFVFWVWSA 382

Query: 418  VTCFGNIFVICMRPYIRSENKLYAMSIIISLCCADCLMGIYLFVIGGFDLKFRGEYNKHAQ 477
          VTCFGNIFVICMRPYIRSENKLYAMSIIISLCCADCLMGIYLFVIGGFDLKFRGEYNKHAQ
Sbjct: 383  VTCFGNIFVICMRPYIRSENKLYAMSIIISLCCADCLMGIYLFVIGGFDLKFRGEYNKHAQ 442

Query: 478  LWMESTHCQLVGSLAILSTEVSLLLLTFLTLEKYICIVYPFRCVVRPGKCRITITVLILIWI 537
          LWMESTHCQLVGSLAILSTEVSLLLLTFLTLEKYICIVYPFRCVVRPGKCRITITVLILIWI
Sbjct: 443  LWMESTHCQLVGSLAILSTEVSLLLLTFLTLEKYICIVYPFRCVVRPGKCRITITVLILIWI 502

Query: 538  TGFIVAFIPLSNKEFFKNYYGTNGVCFPLHSEDTE SIGAQIYSVAIFLGINLAAFIIVF 597
          TGFIVAFIPLSNKEFFKNYYGTNGVCFPLHSEDTE SIGAQIYSVAIFLGINLAAFIIVF
Sbjct: 503  TGFIVAFIPLSNKEFFKNYYGTNGVCFPLHSEDTE SIGAQIYSVAIFLGINLAAFIIVF 562

Query: 598  SYGSMFYSVHQSAITATEIRNQVKKEMILAKRFFFIIVFTDALCWIPFVVKFSLSLQVEI 657
          SYGSMFYSVHQSAITATEIRNQVKKEMILAKRFFFIIVFTDALCWIPFVVKFSLSLQVEI
Sbjct: 563  SYGSMFYSVHQSAITATEIRNQVKKEMILAKRFFFIIVFTDALCWIPFVVKFSLSLQVEI 622

Query: 658  PGTITSWVVFILPINSALNPILYTLTTRPFKEMIHRFWYNYRQRKSMSDSKGQKTYAPSF 717
          PGTITSWVVFILPINSALNPILYTLTTRPFKEMIHRFWYNYRQRKSMSDSKGQKTYAPSF
Sbjct: 623  PGTITSWVVFILPINSALNPILYTLTTRPFKEMIHRFWYNYRQRKSMSDSKGQKTYAPSF 682

Query: 718  IWVEMWPLQEMPPELMKPDLTYPCEMSLISQSTRLNSYS 757
          IWVEMWPLQEMPPELMKPDLTYPCEMSLISQSTRLNSYS
Sbjct: 683  IWVEMWPLQEMPPELMKPDLTYPCEMSLISQSTRLNSYS 722

```

FIG. 5

Signal peptide

N-flank cysteine-rich sequence

Leucine-rich repeats

| | | | | | | | | | | | | | | |
|------|------|----|-----|----|----|------|--------|-------|------------|-----------|-----|------|----|-----|
| LGR4 | RTV- | SE | IHG | SA | QS | RLDA | H- | TSV | EDS-- | FEGLVQLRH | WLD | S-L- | EV | VR |
| LGR5 | RHV- | TE | LQN | RS | QS | RLDA | H- | SYV | P-SC- | FSGLHSLRH | WLD | A-L- | E | VQ |
| LHR | RYIE | -G | FIN | PG | KY | SIC- | TG | RKF | DVTKVFSSES | NFI- | EIC | LHI- | T | GN |
| FSHR | LYIN | -E | FQN | PN | QY | LIS- | TG | KHL | DVHK- | IHSLOKVL- | DIQ | INI- | H- | ERN |
| TSHR | TYID | -D | LKE | PL | KF | GIF- | TGLKMF | DLTK- | VYSTDI | FFI | EIT | PYM- | S | VN |

| | | | | | | | | | | |
|------|---|------------|----------|-------|---------------|--------|-------|------|----|-------|
| LGR4 | T | LNYYNYLDEF | Q-AIKA | PS | KELGFHSNSISVI | D-GA | GGNPL | RTIH | - | DNPLS |
| LGR5 | T | LNYYNNLDEF | T-AIRT | SN | KELGFHSNNIRSI | E-KA | VGNPS | ITIH | F- | DNPIQ |
| LHR | T | ISSTKLQAL | SYGLESIQ | I-ATS | -SYSLKKL | SRET | V-N-- | LEAT | T | ---- |
| FSHR | I | ISRTRIHS | SYGLEN | KK | R-ARSTYN-LKKL | TLEKLV | A-- | MEAS | T | ---- |
| TSHR | L | VSOTSVTAL | SKGLEH | KE | I-ARNTWT-LKKL | LSLS | LH-- | TRAD | S | ---- |

| | |
|------|--|
| LGR4 | FVGNSAFHNLSDLHCLVIRGASLVQWFPNLTGTVHLES�LTGTGKISSIPDDLQCNQKML |
| LGR5 | FVGRSAFQHLPELRTLTNGASQITEFPDLTGANLES�LTGTAQISSLPQTVCNQLPNL |
| LHR | ----- |
| FSHR | ----- |
| TSHR | ----- |

[illegible]

A diagram illustrating the sequence alignment of five G-protein coupled receptors (GPCRs) based on their extracellular domains. At the top, a horizontal line with two arrows pointing left and right indicates a conserved domain. Below this, the amino acid sequences for each receptor are listed and aligned. LGR4 and LGR5 have specific sequences shown, while LHR, FSHR, and TSHR are represented by dashed lines indicating gaps or missing information in the alignment.

| | |
|------|---------------------------------------|
| LGR4 | SGAFAKLGTITNLDVSFNELTSPTEGLNGLNQLK |
| LGR5 | PNAFSTLP SLIKLDLSSNLLSSF PITGLHGLTHLK |
| LHR | ----- |
| FSHR | ----- |
| TSHR | ----- |

| | | | |
|------|--------------------------|------|-----------------------------|
| LGR4 | LVGNFKLKDALAARDFANLRSLSV | YAYQ | WGCDSLCKLNTEDNSPQEHSVTKEKGA |
| LGR5 | LTGNHALQSLISSENFPELKVIM | YAYQ | GVCENAYKISNQWNKGDNSSMDDLHKK |
| LHR | ----- | --SH | RNLPTKEQNFSSHISENFSKQCESTVR |
| FSHR | ----- | --SH | ANWRRQISELHPICNKSILRQEVDMYK |
| TSHR | ----- | --SH | KNOKKIRGILESMLCNESMOSLRORT |

| | | |
|------|---------------|---|
| LGR4 | TDAANVTSTAENE | HS----- |
| LGR5 | DAGMFQAQDERDL | DF----- |
| LHR | KVSNKTLYSSMLA | SE----- |
| FSHR | QTRGQRSSLAEDN | SS----- |
| TSHR | SVNALNSPLHOEY | ENLGDSIVGYKEKSKFODTHNNAHYYVFEEEOEDEIIIGFGQELKNP |

| | | | | | |
|------|--------------------------|---|------|---|----------|
| LGR4 | -----QIIIH | T | STGA | K | YLLGSWMI |
| LGR5 | -----LLDFEEDLKALHSVQ | S | SPGP | K | HLLDGWLI |
| LHR | -----LSGWDEYEGFCLPKTPR | A | EPDA | N | DIMGYDFL |
| FSHR | YSRGFDMTYTEFDYDLNCEVDVT | S | KPDA | N | DIMGYNIL |
| TSHR | QEETLQAFDSHYDYTCIGDSEDMV | T | KSDE | N | DIMGYKFL |

TM 1

| | | | | | | | | | | |
|------|-----|------|-----|-----|-------|--------|-----------|-------------------|-------------|---------------|
| LGR4 | LTV | F | FLV | LLF | LL | ILTVFA | CSS | PASKLFIGLISVSNLLM | IYTGILTFL | AVSW |
| LGR5 | IGV | T | AV | LTC | AL | TSTVFR | PLYISPIKL | IGVIAAVNMLT | VSSAVL | G AF F |
| LHR | VLI | L | NI | IMG | MT | LFVLLT | RYK | TVPRF | MCNLSFADFCM | LYLLLI S SQ K |
| FSHR | VLI | F | SI | ITG | II | LVILTT | QYK | TVPRF | MCNLAFADLCI | IYLLLI S IH K |
| TSHR | IVV | FVSL | LLG | VF | LLLLT | HYK | NVPRF | MCNLAFADFCM | MYLLLI S | LY H |

TM 2

TM 3

| | | | | | | | | | | | | | |
|------|-----------|----|---|-----|-----|-----|---|----|----|-----|----|-----------------|--------|
| LGR4 | GRFAEFG | W | E | S | KV | SLA | S | SA | FL | LA | AV | SVFAKDLMKHGKSSH | QF |
| LGR5 | GSFARHGAW | EN | V | HVI | LSI | | | S | FL | LAA | | GFSVKYSAKFET | APFSSL |
| LHR | GQYYNHA | D | Q | S | ST | FT | | L | YT | VIT | | WHTITYAIHLDQ | LR |
| FSHR | SQYHNYA | D | Q | A | DA | FT | | L | YT | AIT | | WHTITYAMQLDC | VQ |
| TSHR | SEYYNHA | D | O | P | NT | ET | | L | YT | VIT | | WYAITEFAMRLDR | IR |

FIG. 6 (CONT)

8/8

| | <u>TM 4</u> | | | | <u>TM 5</u> | | | |
|------|-------------------|--------|--------|-------------------|-------------|------|------|--|
| LGR4 | QVAALLALLGAAVAGCF | FHGGQ | SASPL | FPTGETPSLGFTVTLVL | SL | LLMA | | |
| LGR5 | KVIILLCALLALTM AV | L G K | GASPL | LPFGEPSTMG | MVALIL | SLC | LMMT | |
| LHR | ILIMLGGWLFSSLI ML | V V N | MKVSIF | MDVETTLISQV | ILILI | VV | FIIC | |
| FSHR | ASVMVMGWIFAFAA LF | IF I S | MKVSIF | MDIDSPLSQL | VMSLLV | VL | VVIC | |
| TSHR | CAIMVGGWVCCFL LL | V I S | AKVSI | MDTETPLALA | IVFVLT | IV | VIVC | |

| | <u>TM 6</u> | | | | | | | |
|------|-------------|--------------------|-------------|--------|----|-------------|---------|------|
| LGR4 | II T L | CNL-EKEDLSENSQSSVI | HV W | NCIFFC | VA | FSFAPLITAIS | SPEI | |
| LGR5 | IA T L | CNL-DKGDLENIW | CSMV HI L L | NCILNC | VA | LSF | SLINLTF | SPEV |
| LHR | AC I I | FAVRNPELMATNK | TKIA KM I | DFTCMA | IS | FAI | AAFKVPL | TVTN |
| FSHR | GC IHI | LTVRNPNIIVSSSS | TRIA RM M | DFLCMA | IS | FAI | ASLKVPL | TVSK |
| TSHR | CCHV I | ITVRNPQYNPGDK | TKIA RM V | DFICMA | IS | YAL | AILNKPL | TVSN |

| | <u>TM 7</u> | | | | | |
|------|-------------|-------|------|-----|-----|----------|
| LGR4 | M | SVTLI | F | LPA | L | V VF N |
| LGR5 | I | FI | LVVV | LPA | L | L IL N |
| LHR | S | VL | VL | Y | INS | A F AI T |
| FSHR | A | IL | VL | H | INS | A F AI T |
| TSHR | S | IL | VL | Y | LNS | A F AI T |

C-terminal tail

| | |
|------|---|
| LGR4 | PK KE WKL KRRVTRKHGSVSVSISSQGGCGEQDFYDCGMYSHLQGNLTVCDCCESFL |
| LGR5 | PH KE LVS RKQTYVWTRSKHPSLMSINSDDVEKQSCDSTQALVTFTSSSITYDLPPSS |
| LHR | KT QR FFL LSKFGCCKRRRAELYRRKDFSAYTSNCKNGFTGSNKPSQSTLKLSTLHCQG |
| FSHR | KN RR FFI LSKCGCYEMQAQIYRTETSSTVHNTHPRNGHCSSAPRVTNGSTYILVPLS |
| TSHR | KA QR VFI LSKFGICKRQAQAYRGQRVPPKNSTDIVQVKVTHDMRQGLHNMEDVYELI |

| | |
|------|---|
| LGR4 | LTKPVSCKHLIKSHSCLPVLTAASCQRPEAYWSDCGTQSAHSDYADEEDSFVSDSSDQVQA |
| LGR5 | VPSPAYPVTESCHLSSVAFVPC |
| LHR | TALLDKTRYTEC |
| FSHR | HLAQN |
| TSHR | ENSHLTPKKQGQISEEYMQTVL |

| | |
|------|---------------------------|
| LGR4 | CGRACFYQSRGFPLVRYAYNLQVRD |
|------|---------------------------|

FIG. 6 (CONT)